Corentin Puffay[1,2], Jonas Vanthornhout[1], Marlies Gillis[1], Bernd Accou[1,2], Hugo Van hamme[2], Tom Francart[1]
[1]KU Leuven, Dept. Neurosciences, ExpORL, Leuven, Belgium, [2]KU Leuven, Dept. of Electrical Engineering, PSI, Leuven, Belgium

Neural tracking of linguistic features at different speech rates using a deep neural network

Background

State-of-the-art (SoA)

- We can measure speech understanding using the neural tracking of linguistic features [4]
- Linear regression models controlling for acoustic information showed that linguistic neural tracking is modulated by speech rate variations [6]
- As a less conservative approach, a multi-input CNN model controlling for phoneme or word onsets showed an added value of linguistic features to the model performance [5]

Aim of the study

Investigate the impact of speech rate on the neural tracking of linguistics over neural tracking of onsets using a nonlinear CNN model.

Dataset (from [6])

- Subjects: EEG of 18 young normal-hearing native Dutch recorded
- Stimulus: one story chunked in non-overlapping segments of variable sizes compressed at six different compression ratios (CRs) between 1.4 and 0.22 to obtain two 5 min fragments per CR and per subject

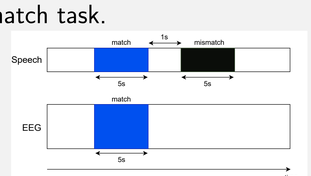
Feature extraction

EEG was processed according to [1], and linguistic and onset features were extracted according to [4].

- Word onsets (WO), phoneme onsets (PO)
- Phoneme-based linguistics: phoneme surprisal (PS), cohort entropy (CE), for a phoneme i: $PS_i = -log_2(\frac{freq(cohort_i)}{freq(cohort_{i-1})})$, $CE_i = -\sum_{word}^{cohort_i} p(word) log_2(p(word))$
- Word-based linguistics: word frequency (WF), word surprisal (WS), for a word i: $WS_i = -log_{10}(p(w_i|w_{i-5}, .., w_{i-1}))$, $WF_i = -log_{10}(p(w_i))$

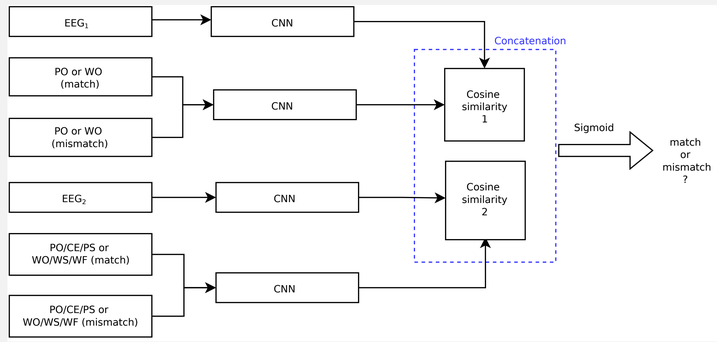
The match-mismatch task [3]

To measure the neural tracking of linguistic and/or lexical features, we use the match-mismatch task.



Multi-input feature model [5]

We investigate the difference between the performance of the **onset** models (2PO or 2WO) and the **linguistics** model (PO1D+CE/PS or WO+WS/WF respectively).



Pre-training, fine-tuning and evaluation

- We pre-train the onset and linguistic models on another dataset (partly published [2]): EEG recordings of 60 young normal-hearing native Dutch subjects listening to 10 stories (14 min each) at a normal speech rate
- We fine-tune the models per CR with half the data of all subjects and evaluate on the other half (90 min each)

Results 1: Word-level linguistic features across speech rates

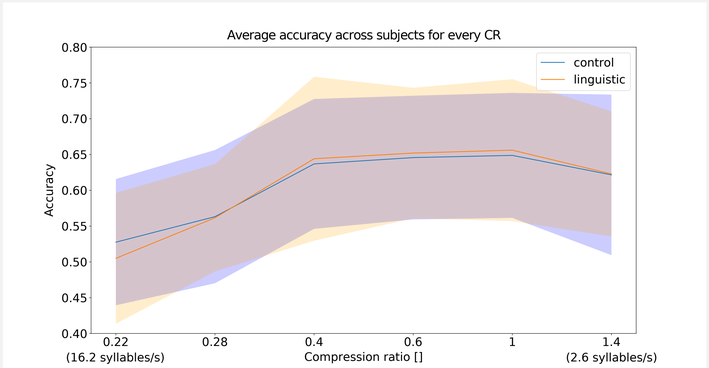


Figure: **Impact of word frequency over word onsets for different compression ratios.** *Difference between control and linguistics model performed with a Wilcoxon signed ranked test, ***: $p<0.001$*

Results 2: Phoneme-level linguistic features across speech rates

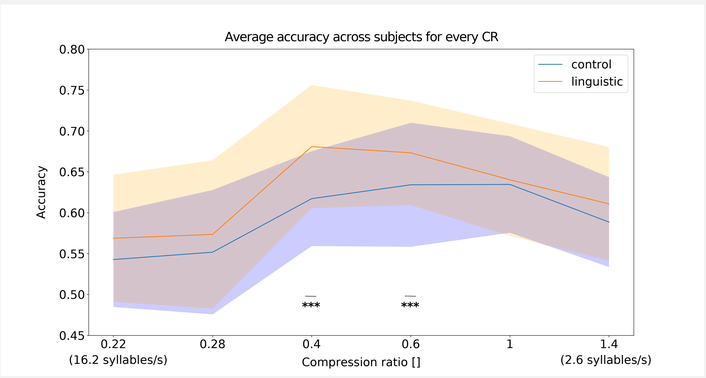


Figure: **Impact of cohort entropy over phoneme onsets for different compression ratios.** *Difference between control and linguistics model performed with a Wilcoxon signed ranked test, ***: $p<0.001$*

Discussion and Conclusion

- Decrease in neural tracking of both onsets and linguistics at high speech rates, which corresponds to behavioral measurements from [6] (CRs of 0.22 and 0.28 had less than 12.5% of speech understanding, from 0.4 and above more than 85%)
- Significant difference in neural tracking between phoneme onsets and cohort entropy at CRs 0.4 and 0.6
- Less conservative control and still an effect of linguistics over onsets at normal speech rates
- Up next: analysis when listening to an unknown language

Acknowledgments

Funding was provided by FWO: FWO fellowships to Bernd Accou (1S89622N), Marlies Gillis (1SA0620N), Corentin Puffay (1S49823N), and Jonas Vanthornhout (1290821N).

Bibliography

[1] B Accou, M Jalilpour-Monesi, J Montoya-Martínez, HV hamme, and T Francart. "Modeling the relationship between acoustic stimulus and EEG with a dilated convolutional neural network". In: *2020 28th European Signal Processing Conference (EUSIPCO)* (2021), pp. 1175–1179.

[2] L Bollens, B Accou, H Van hamme, and T Francart. *A Large Auditory EEG decoding dataset*. Version V1. 2023.

[3] A de Cheveigné, M Slaney, SA Fuglsang, and J Hjortkjaer. "Auditory stimulus-response modeling with a match-mismatch task". In: *Journal of Neural Engineering* 18.4 (2021), p. 046040.

[4] M Gillis, J Vanthornhout, JZ Simon, T Francart, and C Brodbeck. "Neural markers of speech comprehension: measuring EEG tracking of linguistic speech representations, controlling the speech acoustics". In: *Journal of Neuroscience* (2021). eprint: https://www.jneurosci.org/content/early/2021/10/29/JNEUROSCI.0812-21.2021.full.pdf.

[5] C Puffay, J Vanthornhout, M Gillis, B Accou, HV hamme, and T Francart. "Robust neural tracking of linguistic speech representations using a convolutional neural network". In: *Journal of Neural Engineering* 20.4 (2023), p. 046040.

[6] E Verschueren, M Gillis, L Decruy, J Vanthornhout, and T Francart. "Speech Understanding Oppositely Affects Acoustic and Linguistic Neural Tracking in a Speech Rate Manipulation Paradigm". In: *Journal of Neuroscience* 42.39 (2022), pp. 7442–7453. eprint: https://www.jneurosci.org/content/42/39/7442.full.pdf.